



HISTORY ASSOCIATES INCORPORATED

5 CHOKE CHERRY ROAD, SUITE 280
ROCKVILLE, MARYLAND 20850-4004

PHONE: (301) 670-0076 FAX: (301) 670-2765
E-MAIL: HAIGEN@MCIMAIL.COM

July 15, 1997

Margaret Hedstrom
Associate Professor
School of Information
University of Michigan
550 E. University Avenue
Room 304
Ann Arbor, MI 48109-1092

Dear Margaret:

Enclosed is a copy of a report I just completed regarding our meeting in May.
We'll keep you informed of progress on the action agenda developed that day.

Thanks for participating.

Best regards,

James B. Gardner
Project Director

REPORT ON

**DOCUMENTING
THE DIGITAL AGE:
SETTING AN AGENDA
FOR ACTION**

Meeting sponsored by the Commission on Preservation and
Access and History Associates Incorporated

James B. Gardner
Project Director
History Associates Incorporated

INTRODUCTION

With funding support from the Commission on Preservation and Access, History Associates Incorporated (HAI) convened a special meeting on May 23, 1997, to discuss archiving digital information and records. HAI brought together representatives of key agencies and projects to review the status of efforts to archive digital information and records, explore how that work can be better coordinated, discuss funding strategies, and set an agenda for action.

Earlier in the year, History Associates had launched a new initiative on "Documenting the Digital Age." Among the various initiatives focusing on digital archives, HAI's February 1997 conference was unique in its sponsorship (the National Science Foundation, MCI Communications Corporation, and Microsoft Corporation) and in bringing together a wide range of public and private stakeholders, including key figures from the academic, archival, corporate, government, legal, and technology communities. What was common to that conference and other such efforts was a concern about the absence of direction and coordination on this issue. At the February meeting and in subsequent discussions with other interested individuals and agencies, HAI was struck by the range of activities focused on digital archives and information, including conferences and other forums, research and development initiatives, pilot and demonstration projects, and other efforts to develop standards and policies and identify good practice. Reports include *Research Issues in Electronic Records*, published in 1991 by the National Historical Publications and Records Commission; *Preserving Digital Information: Report of the Task Force on Archiving Digital Information*, published in 1996 by the Commission on Preservation and Access and the Research Libraries Group; *Electronic Records Research and Development*, a report published in 1997 regarding a conference held at the University of Michigan, Ann Arbor in 1996; and *Documenting the Digital Age*, the forthcoming report from HAI's February conference. While these and other publications provide valuable analysis and recommendations, they also testify to a lack of coordination among the various players and funding agencies and the failure to develop a coherent national agenda. The May meeting was convened to address this need.

In consultation with Deanna B. Marcum, president of the Commission on Preservation and Access and the Council on Library Resources, and Philip L. Cantelon, president of History Associates, project director James B. Gardner developed the meeting agenda and invited individuals to participate. For a copy of the meeting agenda and a list of participants, see appendices A and B respectively. Although the meeting was arranged on very short notice, HAI was able to recruit a mix of participants from the federal government, nonprofits, and corporations, including specialists in archives and computer science, authors of reports on electronic archives issues, and representatives of funding

agencies, related organizations and coalitions, and special projects. Prior to the meeting, each participant was sent a copy of the four reports noted above plus related material provided by the Research Libraries Group.

The following report summarizes the discussion, organized according to the day's agenda.

OPENING DISCUSSION

Philip L. Cantelon, History Associates Incorporated, chaired the meeting. After introductions around the table, Cantelon reviewed the goals of the meeting--to move beyond the usual discussions within the archives and library communities and establish the basis for new partnerships among the various interest groups with a stake in the preservation of electronic information and records. He urged the participants to "think outside the box" and consider other kinds of records and other groups to work with. He then asked for oral summaries of the four key reports.

Gerald George, National Historical Publications and Records Commission (NHPRC), spoke briefly about the report published by his agency in 1991 regarding the recommendations of the Working Group on Research Issues in Electronic Records, a group brought together by the Minnesota Historical Society with funding from NHPRC. The group identified ten questions on which research needed to be done, and NHPRC has since given priority to the first three in its funding of electronic records research. To assess progress toward answering those questions, representatives of NHPRC projects were invited to participate in a conference in 1996 on Electronic Records Research and Development, held at the University of Michigan, ANN Arbor (see below).

Donald Waters, Yale University, summed up the work of the Task Force on Digital Archiving, established by the Commission on Preservation and Access and the Research Libraries Group and cochaired by him. He called the group's attention to the report's Executive Summary and reviewed the Task Force's work--its delineation of the features that determine information integrity and of the functional activities of digital archives, its shift from refreshing to migration as a central preservation strategy, its concerns about establishing a certification process for digital repositories and a "fail-safe mechanism," and its recommendations re developing pilot projects, establishing support structures, and identifying best practices. He noted that the Task Force had begun with the assumption of the need to preserve electronic information but suggested that we cannot assume the value of such information, that we need to look at who needs what and for what purpose.

Margaret Hedstrom, University of Michigan, Ann Arbor, reported on the June 1996 conference at Ann Arbor following up on the research agenda established by NHPRC in 1991. The purpose of the conference was to assess progress on that agenda, looking at what has changed in the external environment and how to improve the way research is carried out. She referred to the report's Executive Summary for the conference's findings and recommendations for new directions and priorities in research and development. She noted progress in designing standards for systems to create and

maintain usable electronic records but expressed frustration in getting standards adopted early enough in systems design, which may be due in part to archivists' lack of knowledge about the environment outside government and universities. She emphasized the need to connect with compatible research projects within the digital libraries area and urged more collaboration with the private sector.

James B. Gardner reviewed the work of the conference convened by HAI in February 1997. He emphasized HAI's commitment to broadening the discussion beyond the archival and library communities and noted the range of interests represented in the conference. Referring to the "Findings" section of the draft conference report, he noted the persistent tension between what archivists argue should be done and what technology specialists claim could be done, between archival principles and the technological capability to save everything. He emphasized the importance of preserving corporate and private as well as public information and records and suggested that the real obstacles are legal rather than financial or technological. Finally, he noted the concerns of conference participants regarding maintaining the integrity and functionality of digital material.

When Cantelon opened the meeting to discussion, the group focused primarily on the legal issues raised by the HAI conference and the problem of determining whether electronic communication is a record. In regard to legal issues, the main concern was the issue of discoverability and its impact on corporate record keeping. Cantelon argued that we must recognize that the specter of litigation leads many corporations to purge rather than save electronic records and information, retaining only what is required for legal and financial reasons. He referred participants to an article in *The Wall Street Journal* (May 22, 1997) on the "culling" of corporate files and an op-ed from *The Washington Post* (May 16, 1997) focusing on the impact of discoverability on note-taking and memo-writing within the federal government. Jeff Rothenberg, the President's Committee on the Arts and Humanities, noted that it is not only the expense of the discovery process that corporations want to avoid but also the possibility that the process may uncover otherwise unknown records, which in turn become the basis for further suits. Hedstrom suggested that electronic records management can provide some remedy by making it possible to segregate otherwise intermingled records and better assess costs, benefits, and risks. But Gardner reminded the group that many corporations are not interested in making discovery less difficult; instead, their goal is to make sure that there is nothing there to be discovered. Waters noted that that problem applies to paper as well as digitized records. Cantelon emphasized again that corporations see all costs and no benefits in preserving anything other than what is required for legal and financial reasons and that the electronic archives community must argue that addressing records needs on the front end is less costly than not.

Rothenberg suggested that part of the problem is the lack of clarity in defining what role electronic communication plays in record keeping: is e-mail analogous to the informal communication of a phone conversation, to the more formal role of a letter, or to some intermediate between the two? How we view electronic communication has implications for whether we save it or not. Participants discussed the use of electronic communication for policy, procedural, and other institutional matters; for financial and legal record keeping; and in electronic commerce, including contracts. Louis I. Reich, Computer Science Corporation, noted that in many cases the online copy--not the paper one--is considered the official record.

The group then focused its attention on the questions raised in the meeting agenda.

WHAT ARE OUR GOALS? WHAT ARE THE AREAS OF COMMON INTEREST?

The group came up with six basic goals, each of which is discussed below along with a summary of discussion points.

1. **Develop better connections and more collaborations among the various communities that have a stake in digital archiving.** The group agreed that work thus far has focused too narrowly on needs and concerns within the government, archives/library, and research communities and should broaden to encompass business, computer science and technology, museums and other cultural organizations, science, and other areas that have a stake in digital archiving. The purpose of such cross referencing should be to identify key issues and develop broader collaboration. Particular effort should be made to bring foundations, other government agencies, and the private sector into collaborative projects.
2. **Identify needs and incentives for digital archiving.** Outside the archival community, there is little understanding of what archiving means, of the difference between saving and archiving records and information. Progress in digital archiving will depend on clarifying goals and developing compelling arguments that address needs, identify incentives, and provide concrete models for good documentation and record keeping over time.
3. **Identify legal issues in tension with digital archiving.** The group identified intellectual property issues and problems related to discoverability as the two areas of concern. Resolution of these problems will depend on the digital community's willingness to work collaboratively toward appropriate changes in law and policy.
4. **Develop systems that will be simple and easy to use and will preserve functionality.** All agreed on the necessity of developing simple and unobtrusive tools to facilitate digital archiving. If difficult or clumsy to use, archiving functions will not be implemented. While some advocated embedding such tools in software, others argued for focusing on the backup process. In any case, archiving must become an integral part of the work process, a natural byproduct. Systems must also address concerns about preserving functionality. While some participants advocated pursuing migration strategies, others argued that migration is premature and problematic and that the wisest course is to preserve software and hardware emulations. All agreed, however, that functionality is essential to context and must be preserved as part of the record.

5. **Develop the underlying technology for preservation.** Computer science must address how to preserve digital records and information for the long term.
6. **Move forward with developing both concrete examples and broader standards and frameworks.** Participants agreed that both are essential to building workable solutions but that there must be better coordination and communication--we need to build on experience not keep reinventing. One solution would be a clearinghouse, although there was disagreement over the feasibility of such an effort.

WHAT BARRIERS DO WE FACE?

The group found four barriers, each of which is discussed below along with a summary of discussion points.

1. **Lack of understanding of the problem outside the archival community.** Participants expressed concern that digital preservation is seen as a problem only within the archival community. Within the corporate and computer science communities, for example, the concerns are access and dissemination not preservation. The archival community needs to make a stronger and clearer case for preservation, identifying and responding to more immediate needs and incentives within the marketplace. On the other hand, no archiving system is likely to provide the full contextualization that historians expect. The archival community needs to make clear what it can and cannot do.
2. **Uncertainty within the archival community about what it is we are trying to preserve.** What is the goal of digital archiving? When archivists talk about preserving functionality, it is not clear what part of the technology should be preserved as part of the record. For example, in preserving e-mail do we actually need to preserve full functionality? What is essential? Progress cannot be made without clarifying such issues.
3. **Lack of interest on the part of the computer science community.** At this point, many in the computer science community are not interested in preservation technology issues and thus are not developing the technological solutions necessary to move forward. They are more interested in moving into the future, not preserving the past. But there is a common interest that might be used to attract computer scientists--the latter share archivists' concerns about the problem of saving digital information in ways that it can be used in the future, and they might be persuaded to make temporal interoperability a research and development priority. Although participants discussed working with information science organizations to change the mind set of the scientific community, all agreed that the key to changing research priorities is to change the funding priorities of ARPA and NSF's Digital Libraries Initiative.
4. **Legal obstacles.** As noted in the previous sections of this report, there are numerous legal impediments to digital preservation--what can be archived is restricted by intellectual property issues, and concerns about discoverability and liability lead to the purging rather than archiving of records.

WHAT SHOULD OUR PRIORITIES BE-- ARCHIVAL, LEGAL, AND TECHNICAL? WHERE DO WE DIFFER?

The group identified five priorities, each of which is discussed below along with a summary of discussion points.

1. **Bring in other communities.** The archival community should develop partnerships and other collaborative relationships with corporations, government agencies, computer science and technology, and other interested groups--the larger context for digital archiving. Through such links, the archival community can learn more about how electronic information is being managed now and identify needs and incentives for archiving. The archival community should also look for ways to engage foundations and other funders and thereby broaden the financial base.
2. **Work with computer science and technology specialists to develop the underlying technology for preservation.** There will be no solution until the basic technology is in place.
3. **Develop concrete, practical solutions.** The archival community should move beyond narrowly focused research and development projects and develop both demonstration projects that can be replicated in other contexts and prototypes that can be scaled up to address larger needs. In both cases, the goal should be to address actual needs with as little obtrusion as possible.
4. **Disseminate technology and knowledge.** The archival community should consider ways to market digital archiving. This should include fostering broader understanding of preservation needs, establishing a clearinghouse or some other mechanism to facilitate communication among digital archive projects, disseminating system specifications to encourage commercial development, and marketing successful preservation solutions.
5. **Work to resolve legal obstacles.** As noted in previous sections, efforts to move forward with digital archiving will languish without changes in policies and laws regarding intellectual property and discoverability.

HOW DO WE COORDINATE OUR WORK AND MOVE FORWARD?

Finally, the group agreed on specific actions to take, each of which is discussed below along with a summary of discussion points.

1. **Develop communities of interest within the corporate sector.** The participants agreed on two courses of action to establish the basis for partnerships and other collaborations. First of all, Cantelon will contact Robert F. Williams, Cohasset Associates, re the possibility of using his firm's "Managing Electronic Records" conference to engage more corporate decision makers. The second strategy is to repackage digital archiving to address the different objectives of the corporate world. This should be part of a broader marketing and dissemination effort to build support for digital archiving in government, computer science and technology, and other communities. The archival community must recognize, however, that corporations are likely to resist the involvement of outside groups in internal matters.
2. **Coordinate funding.** HAI will pursue further the possibility of bringing together representatives of foundations, government agencies, and other funders to coordinate funding the various pieces needed to make the whole.
3. **Mobilize the computer science community to focus on preservation technology.** Rothenberg volunteered to see what he could do re the key professional organizations, and Hedstrom agreed to raise the issue with NSF.
4. **Begin effort to address legal issues.** HAI will explore possible strategies for developing coalitions or collaborative action to resolve intellectual property and discoverability problems.

APPENDIX A MEETING AGENDA

9:00 a.m.	Continental breakfast
9:00-9:30 a.m.	Introductions
9:30-10:45 a.m.	What are our goals? What are the areas of common interest?
10:45-11:00 a.m.	Break
11:00-12:00 p.m.	What barriers do we face?
12:00-1:00 p.m.	Lunch (provided)
1:00-2:00 p.m.	What should our priorities be--archival, legal, and technical? Where do we differ?
2:00-3:00 p.m.	How do we coordinate our work and move forward?
3:00 p.m.	Adjournment

APPENDIX B LIST OF PARTICIPANTS AND AFFILIATIONS

Philip L. Cantelon
President
History Associates Incorporated
5 Choke Cherry Road
Suite 280
Rockville, MD 20850-4004
Phone: 301/670-0076
Fax: 301/670-2765
E-mail: pcantelon@mcimail.com

Mary Case
Association of Research Libraries
and the Coalition for Networked
Information
21 Dupont Circle
Washington, DC 20036
Phone: 202/296-2296, ext. 112

Alf Erlandsson
7000 Kenhill Road
Bethesda, MD 20817
Phone: 301/229-3981
Fax: 301/229-6928
E-mail: aerlandsson@worldnet.att.net

Jeffrey Field
Deputy Director
Division of Preservation and Access
National Endowment for the Humanities
1100 Pennsylvania Avenue NW
Room 411
Washington, DC 20506
Phone: 202/606-8570
Fax: 202/606/8639
E-mail: preservation@neh.fed.us

Gerald George
Executive Director
NHPRC
National Archives Building (Archives I)
Room 607
Washington, DC 20408
Phone: 202/501-5610
Fax: 202/501-5601

James B. Gardner
Project Director
Documenting the Digital Age
4000 Massachusetts Avenue NW
#228
Washington, DC 20016-5108
Phone: 202/363-3420
Fax: 202/785-3948
E-mail: jbgardner@compuserve.com

David L. Green
 Director
 National Initiative for a Networked
 Cultural Heritage
 21 Dupont Circle
 Washington, DC 20036
 Phone: 202/296-5346
 Fax: 202/296-5436

Margaret Hedstrom
 Associate Professor
 School of Information
 University of Michigan
 550 E. University Avenue
 Room 304
 Ann Arbor, MI 48109-1092
 Phone: 313/747-3582
 Fax: 313/764-2475
 E-mail: hedstrom@umich.edu

Heather Heywood
 World Bank Group Archives
 Room HB1-001
 1818 H Street NW
 Washington, DC 20433
 Phone: 202/473-9453
 Fax: 202/477-1499
 E-mail: hheywood@worldbank.org

Gerald Kovach
 Senior Vice President, Government
 Relations, Law, and Public Policy
 MCI Communications Corporation
 1801 Pennsylvania Avenue NW
 Washington, DC 20006
 Phone: 202/887-2754

Deanna B. Marcum
 President
 The Commission on Preservation and
 Access
 The Council on Library Resources
 1400 16th Street, NW
 Suite 715
 Washington, DC 20036-2217
 Phone: 202/939-3400
 Fax: 202/939-3407

James P. Michalko
 President
 Research Libraries Group
 1200 Villa Street
 Mountain View, CA 94041-1100
 Phone: 415/962-9951
 Fax: 415/964-0943

Louis I. Reich
 Computer Sciences Corporation
 10110 Aerospace Road
 Lanham/Seabrook, MD 20706
 Phone: 301/794-1859
 Fax: 301/794-2280
 E-mail: louis.i.reich@gsfc.nasa.gov

Jeff Rothenberg
 Technology Consultant
 President's Committee on the Arts and
 Humanities
 1609 Hill Street
 Santa Monica, CA 90405
 Phone: 310/450-3922
 E-mail: jeff_rothenberg@acm.org

Kenneth F. Thibodeau
Director
Center for Electronic Records
National Archives and Records
Administration
8601 Adelphi Road
College Park, MD 20740-6001
Phone: 301/713-6630
Fax: 301/713-6911
E-mail: ken.thibodeau@arch2.nara.gov

Donald Waters
Associate University Librarian
Yale University
116B
Sterling Memorial Library
P.O. Box 208240
New Haven, CT 06520
Phone: 203/432-4889
Fax: 203/432-7231
E-mail: donald.waters@yale.edu